Quelle régulation pour les deepfakes de type “Amandine Le Pen” ?

Philippe Mouron

Quelle régulation pour les *deepfakes* de type “Amandine Le Pen” ?

-

***Les Surligneurs*, 18 avril 2024**

Philippe Mouron
Maître de conférences HDR en droit privé
Directeur du Master 2 Droit des médias électroniques – Université d’Aix-Marseille

Le droit actuel n’envisage que de façon imparfaite les *deepfakes* tel quel ceux des soi-disant nièces de Marine Le Pen qui appellent à voter à l’extrême droite aux élections européennes. Le droit européen n’est pas très précis non plus, d’où une proposition de loi en cours de discussion au Parlement. Mais il n’est pas sûr qu’elle suffise non plus…

****

Tantôt « prête à défendre la France », tantôt « trop occupée à bronzée pour me présenter aux prochaines élections » (les fautes d’orthographe ont été conservées), Amandine Le Pen, présentée comme l’une des nièces de Marine, inonde les réseaux sociaux de vidéos dans lesquelles elle joue de sa troublante ressemblance avec sa tante. Elle a été en cela rejointe par Chloé Le Pen et Léna Maréchal, autres membres de la famille, et proclamées influenceuses d’extrême droite.

Seul *hic*, ces trois personnes n’existent pas et sont des *deepfakes* réalisés par intelligence artificielle. Il n’empêche qu’elles génèrent déjà des dizaines de milliers de vues sur TikTok ou X, et sont suivies par de très nombreux internautes, sidérés ou amusés, certains croyant avoir affaire à d’authentiques influenceuses. Le sujet inquiète déjà depuis quelques années, alors que le procédé est de plus en plus utilisé pour détourner l’image et la voix de véritables personnalités, les progrès accomplis en la matière rendant difficile l’identification de ces contenus. Il est d’autant plus d’actualité que commence la campagne des prochaines élections européennes, à laquelle font déjà référence les trois fausses nièces de Marine Le Pen.

Jordan Bardella s’est lui-même ému de la recrudescence de ces vidéos, appelant à une régulation pour éviter de « bousculer le fonctionnement de la démocratie », alors qu’un nouveau délit est en passe d’être adopté dans le projet de loi visant à sécuriser et réguler l’espace numérique (SREN).

L’insuffisance des textes existants

Les moyens de sanctionner la diffusion de *deepfakes* sont en vérité multiples, mais tous imparfaits.

L’article 9 du code civil, qui prévoit que « *Chacun a droit au respect de sa vie privée* », peut être mobilisé en premier lieu dès lors qu’un tel contenu procède de l’image et de la voix d’une personne identifiable. Le retrait d’images fixes ou de vidéos peut être ordonné par les tribunaux sur ce fondement : toute personne dispose sur son image et sa voix d’un droit exclusif, lui permettant de s’opposer à sa reproduction sans son autorisation. Ce droit connaît malgré tout des limites concernant les personnalités publiques, ce sur quoi nous reviendrons.

En droit plusieurs articles du code pénal incriminent la captation et la diffusion d’images et de paroles de personnes dans certaines circonstances. L’article 226-1 punit d’un an d’emprisonnement et de 45 000€ d’amende de tels actes lorsqu’ils sont de nature à porter atteinte à l’intimité de la vie privée ; l’article 226-2-1 sanctionne de deux ans d’emprisonnement et de 60 000€ d’amende la diffusion de telles images lorsque, en plus, elles présentent un caractère sexuel. Mais c’est surtout l’article 226-8 qui semble le plus adéquat, punissant d’un an d’emprisonnement et de 15 000 € d’amende « *le montage réalisé avec les paroles ou l’image d’une personne sans son consentement, s’il n’apparaît pas à l’évidence qu’il s’agit d’un montage ou s’il n’en est pas expressément fait mention* ».

Si les *deepfakes* apparaissent à première vue comme des montages au sens de cette disposition, on doit rappeler que le principe d’interprétation stricte de la loi pénale impose une lecture qui peut en dérouter plus d’un. La notion de montage s’entend classiquement comme l’opération portant sur des images préexistantes ou paroles enregistrées dont la manipulation, notamment la coupe, permet de créer une nouvelle image d’une personne dans une situation fausse ou erronée. Autrement dit, il n’est pas sûr que l’article 226-8 puisse appréhender toutes les variantes de *deepfakes*. Si ceux-ci sont bien réalisés à partir d’images et de sons existants, ce n’est qu’au stade de l’entraînement des systèmes ayant permis de les générer : le résultat visible par l’internaute peut être entièrement nouveau, voire représenter une personne imaginaire. Il en est de même des autres textes précités, qui ne visent que la captation, l’enregistrement ou la transmission de l’image ou des paroles d’une personne réalisés à son insu.

Une répression relative des *deepfakes* dans le projet de loi SREN

Le projet de loi SREN, dont la dernière version a été adoptée en commission mixte paritaire le 10 avril, entend justement renforcer la répression des *deepfakes* sur le terrain pénal.

L’alinéa premier de l’article 226-8 précité du code pénal serait ainsi complété d’une nouvelle phrase : « *Est assimilé à l’infraction mentionnée au présent alinéa et puni des mêmes peines le fait de porter à la connaissance du public ou d’un tiers, par quelque voie que ce soit, un contenu visuel ou sonore généré par un traitement algorithmique et représentant l’image ou les paroles d’une personne, sans son consentement, s’il n’apparaît pas à l’évidence qu’il s’agit d’un contenu généré algorithmiquement ou s’il n’en est pas expressément fait mention* ». Les peines

seraient portées à deux ans d’emprisonnement et 45 000€ d’amende lorsque le délit est commis par la voie d’un service de communication au public en ligne.

La répression de ces pratiques pourrait donc voir le jour plus vite qu’il n’y paraît… mais elle restera insuffisante contre la diffusion de *deepfakes* de personnes imaginaires ! Comme indiqué dans le texte du projet de loi, le nouveau délit ne permettra de sanctionner que les contenus relatifs à une personne identifiée ou au moins identifiable. Cela n’est pas négligeable dans le contexte actuel de *deepfakes* mettant en scène des personnalités politiques. Mais cette infraction nouvelle, si elle est adoptée par le Parlement, n’a pas vocation à jouer lorsque l’on a affaire à de pures créations, comme c’est le cas avec prétendues nièces de Marine Le Pen, qui sont des personnages de fiction.

De surcroît, le délit ne serait constitué qu’en présence d’une volonté de tromper le public, soit parce que le *deepfake* est réaliste, soit parce qu’il n’est pas identifié – ou étiqueté – comme ayant été généré par intelligence artificielle. Or les comptes à l’origine des vidéos en question comportent bien une mention selon laquelle elles sont générées par intelligence artificielle. Celle-ci n’est certes plus visible lorsque les vidéos sont partagées par d’autres utilisateurs, mais le problème tient plus à la modération et au référencement des contenus par les plateformes.

Deepfakes et droit européen

L’article 35 du règlement sur les services numériques (ou *Digital Services Act*) impose aux très grandes plateformes et aux très grands moteurs de recherche en ligne de prendre des mesures d’atténuation des risques systémiques, notamment avec « *le recours à un marquage bien visible pour garantir qu’un élément d’information, qu’il s’agisse d’une image, d’un contenu audio ou vidéo généré ou manipulé, qui ressemble nettement à des personnes,* […] *et apparaît à tort aux yeux d’une personne comme authentique ou digne de foi, est reconnaissable lorsqu’il est présenté sur leurs interfaces en ligne, et, en complément, la mise à disposition d’une fonctionnalité facile d’utilisation permettant aux destinataires du service de signaler ce type d’information* ».

Deepfakes, personnalités publiques, liberté de création et d’expression

Si les hommes et femmes politiques bénéficient comme tout un chacun du droit au respect de leur image, ce droit doit malgré tout être concilié avec l’exercice de la liberté de création. Celui-ci est d’autant plus important lorsqu’il est question de sujets d’intérêt général, et de personnes occupant des fonctions publiques.

Des œuvres de fiction peuvent dès lors être inspirées de ces personnes, l’effet de distanciation pouvant neutraliser un certain nombre d’atteintes à leur image. C’est ainsi que la Cour européenne des droits de l’Homme a jugé disproportionnée l’interdiction d’une œuvre d’art constituée d’un photomontage à caractère sexuel de personnalités politiques autrichiennes, dont notamment Jörg Haider, qui était alors leader d’un parti d’extrême-droite. Quand bien même le tableau avait un caractère outrageant, il n’avait pas vocation à « *refléter ou même évoquer la réalité* » et devait être considéré comme purement satirique. Seules peuvent être sanctionnées

des mises en cause gravissimes, flirtant avec des faits réels, et susceptibles de nuire à la réputation en dépit du caractère fictionnel. Tel a été le cas à propos d’un roman mettant en scène Jean-Marie Le Pen, inspiré de faits authentiques, et qualifiant l’intéressé de « *chef d’une bande de tueurs* » ou encore de « *vampire qui se nourrit de l’aigreur de ses électeurs mais aussi parfois de leur sang* ».

Les *deepfakes* n’échappent pas à ces principes, et ne constituent qu’une nouvelle forme de création qui peuvent parfaitement prendre pour cible des personnes ou situations politiques à des fins satiriques. Les progrès réalisés par l’intelligence artificielle permettent déjà de se délecter des (faux) talents de danseur d’Emmanuel Macron, des (fausses) reprises de Charles Aznavour par Jean-Marie Le Pen ou de celles de « Cheb Marina et Cheb Ricky », en fait Marine Le Pen, Éric Zemmour accompagnés par Gérald Darmanin. Les vidéos d’Amandine Le Pen vont encore bien au-delà puisqu’elles prennent délibérément pour sujet un personnage et une situation fictives.

On peut malgré tout s’interroger, comme l’a fait Jordan Bardella, sur l’éventuelle influence de ces créations sur la campagne électorale à venir. À ce niveau, la régulation ne peut reposer que sur un filtrage ou une modération des contenus partagés sur les réseaux sociaux, tâche qui incombe à leurs seuls fournisseurs. La Commission européenne s’est déjà saisie du problème en publiant fin mars des lignes directrices pour l’atténuation des risques systémiques en ligne pouvant affecter les processus électoraux. L’obligation de marquage des contenus générés par intelligence artificielle, figurant à l’article 35 du DSA, y est précisée ; les plateformes et réseaux sociaux devront ainsi maintenir ce marquage chaque fois qu’un contenu sera partagé par un utilisateur. Autrement dit, le marquage « généré par IA », apposé par le créateur du *deepfake*, ne doit pas disparaître au premier partage par l’utilisateur, et doit au contraire subsister tout au long de la vie du *deekfake* et de sa diffusion par les partages. La Commission a demandé aux principaux fournisseurs de ces services des explications sur les mesures qu’ils entendent prendre pour se conformer à cette obligation.

À voir si cela sera suffisant à court terme.